


Digitaler Datenerhalt: Un projet de l'ETH-Bibliothek pour l'ETH Zurich

Archivage numérique dans le Domaine des EPF 20 mars 2012

Presentation

Author(s):

Töwe, Matthias 

Publication date:

2012

Permanent link:

<https://doi.org/10.3929/ethz-a-007362116>

Rights / license:

In Copyright - Non-Commercial Use Permitted

Digitaler Datenerhalt: Un projet de l'ETH-Bibliothek pour l'ETH Zurich

Archivage numérique dans le Domaine des EPF

20 mars 2012

Dr. Matthias Töwe

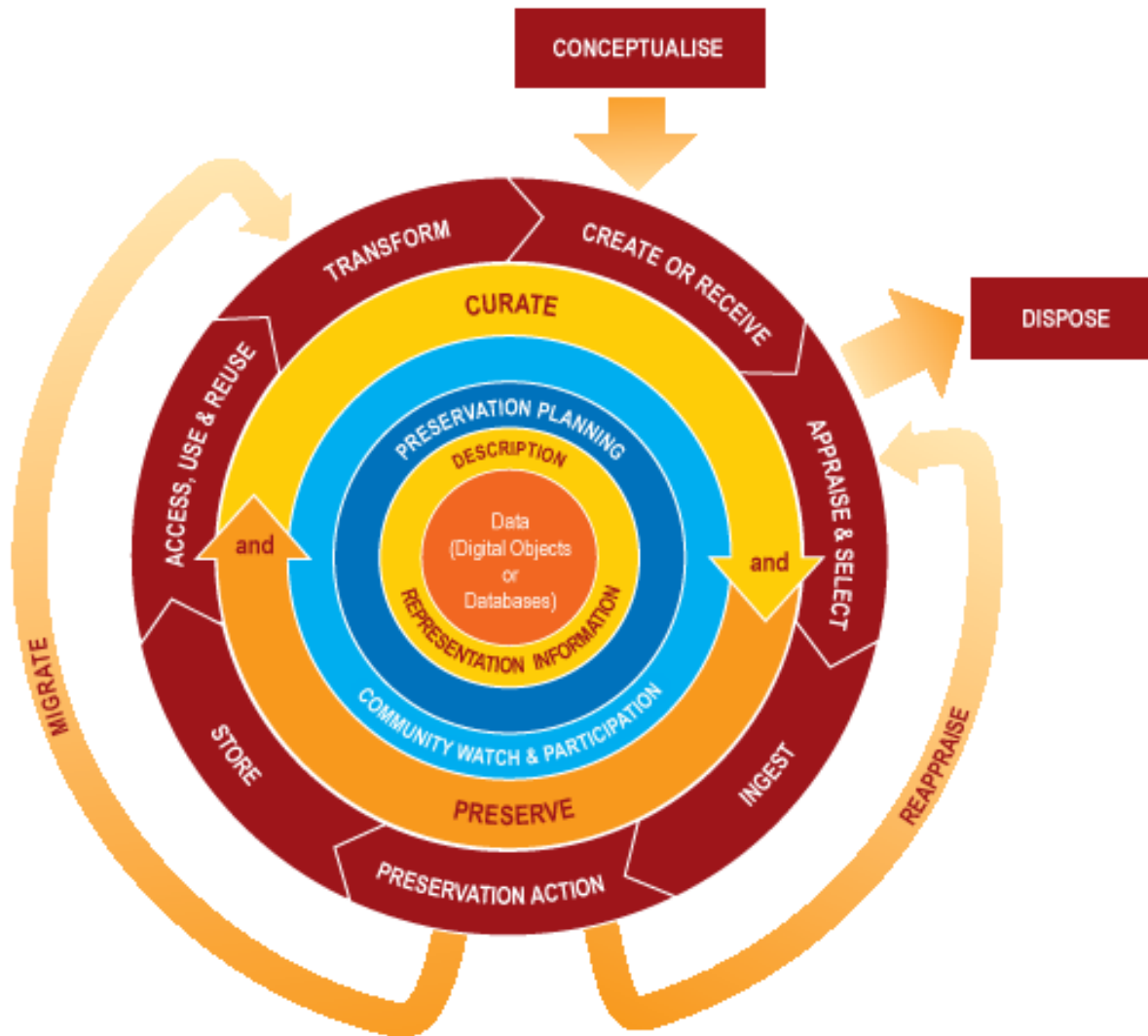
STRUCTURE

1. Situation
2. Objectifs
3. Status quo selon sondage
4. Vision
5. Prochaines étapes

SITUATION

- Majorité des résultats de recherche existent **uniquement sous forme numérique**
- Données des observations et des enquêtes ne sont **pas reproductible**, données du laboratoire pas reproductible en fonction des échantillons et des méthodes
- **Exigences** des hautes écoles et des institutions d'encouragement de la recherche:
 - **Vérification** (cf. ETH- Richtlinien für Integrität in der Forschung)
 - **Plans de gestion des données** dans demandes de la recherche (entre autres NSF, DFG; SNF suivra probablement)
- **Intérêt des éditeurs** pour *Supplementary Material*
- Dossiers administratifs de l'ETH Zurich (**Archives d'ETH Zurich**)
- **Collections** (uniques) de l'**ETH-Bibliothek**

CYCLE DE VIE



The DCC Curation
Lifecycle Model

Digital Curation
Centre, UK
Copyright 2010

OBJECTIFS

- Données pas reproductible (ou avec effort prohibitif) sont disponibles pour l'**utilisation ultérieure**
- Données peuvent être **citées** à manière durable et être **référéés** par publications
→ enregistrement des DOI déjà en service
- **Reproductibilité et vérification par tiers** sont assurées
- Si besoin est, données peuvent être **mises à disposition pour tiers** avec effort insignifiant
- Il n y a **pas une contrainte** de publier des données



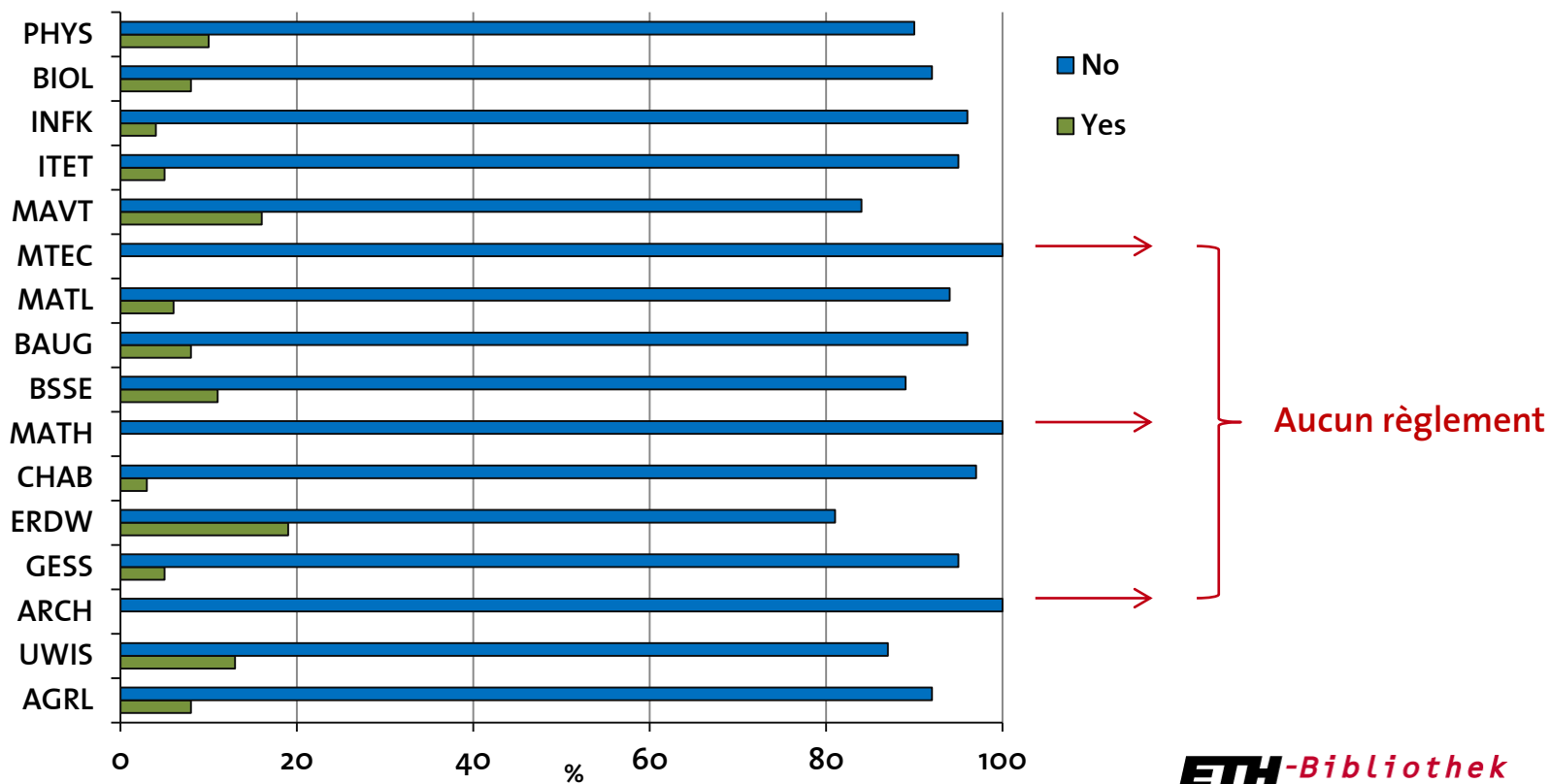
Helping you to find,
access, and reuse data

SONDAGE „MANIEMENT DES DONNÉES DE RECHERCHE“

- **Sondage de tous le 450 groupes de recherche** dans les 16 départements de l'ETH Zurich
- **Sondage en ligne** en accord avec les responsables du support d'IT et les commissions indiquées par eux
- **Rappels** par téléphone et en personne
- **Taux de réponses** pour toute l'ETH:
 - **80%** des groupes de recherche
 - Env. 65% de réponses par professeurs, agrégés, assistants supérieurs

SONDAGE „MANIEMENT DES DONNÉES DE RECHERCHE“

1. *part* – Est-ce que votre institut ou votre groupe a un règlement écrit sur la gestion des données propres ou des données des tiers ou existent des documents similaires, qui décrivent le maniement des données de recherche ?

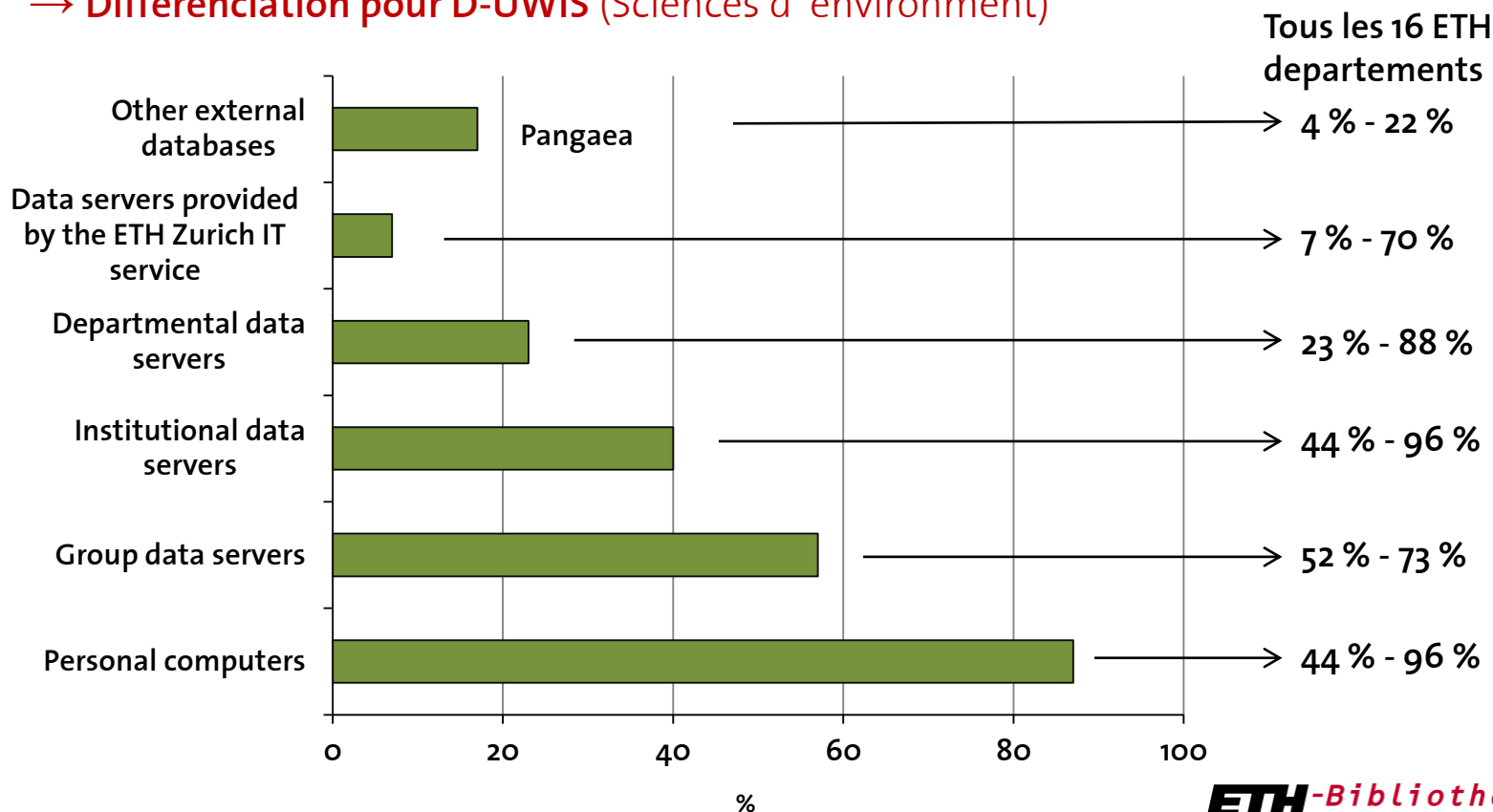


SONDAGE

„MANIEMENT DES DONNÉES DE RECHERCHE“

2. *part* - Quelles possibilités pour le stockage des données des recherche vous ou votre groupe de recherche utilisez ?

→ Différenciation pour D-UWIS (Sciences d' environnement)



RÉSULTAT : VOEUX

Chercheurs...

- voudraient **garder le contrôle** complet sur l'accès.
- voudraient **restructurer et compléter des données** avant l'Ingest, et **éditer les métadonnées**.
- sont intéressées à un **support de la pérennisation et du contrôle de qualité** (check-lists, feed-back sur les métadonnées ...).
- doivent garder beaucoup des données pendant une **période limité** (p. ex. pour 10 à 12 années).
- voient le besoin visant l'archivage souvent en relation avec des **données qui sont utilisées pour des publications**.
- se prononcent en manière différente sur un **règlement des données pour l'ETH**.
- ne voudraient **pas devoir effectuer un effort supplémentaire** sans valeur ajoutée claire.

EXIGENCES

- **Multiplicité** des formats de fichiers
- **Volume** des données
- **Intégration des systems existants** dans le groupe de recherche (interfaces avec instruments de mesure et logiciels)
- Conservation pour la **vérification** (typiquement : 10 années)
- **Pérennisation** des données non reproductibles
- Si besoin est : mise à disposition pour utilisation par tiers (**Open Data**)
- **Documentation** déjà pendant la saisie (metadonnées)
- **Motivation** des chercheurs
- **Effort et dépenses** initiales et permanentes

NIVEAUX FONCTIONNELS

Quoi ?

Data Curation

Content
Preservation

Bitstream
Preservation

Pourquoi ?

Ensure intellectual
re-usability

Ensure technical
re-usability

Ensure technical
stability

Qui ?

Data Producers

ETH-Bibliothek
(Rosetta)

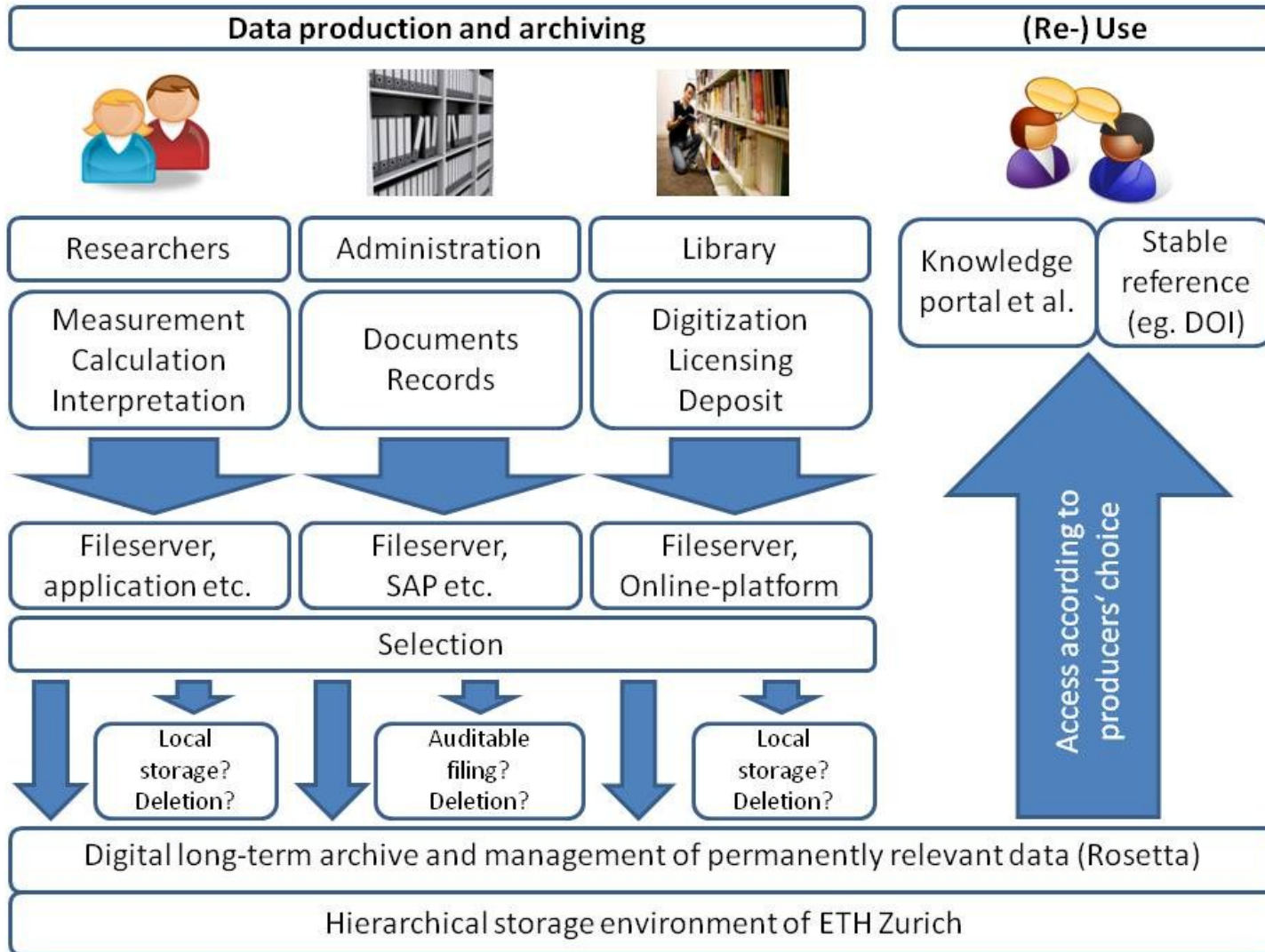
IT-Services
ETH Zurich

Adapted after Jens Ludwig, Wissgrid

DIFFÉRENCES ENTRE TYPES DES DONNÉES ?

Quoi ?	Données de recherche	Objets de la bibliothèque
Data Curation	Comprehensive documentation by producers required	Full control of metadata and context
Content Preservation	More and less common formats	Mainly standard formats
	Same preservation procedures apply	
Bitstream Preservation	„Any object is just bits“	

VISION – VUE D'UTILISATEUR



Une
base
technique
pour
services
divers

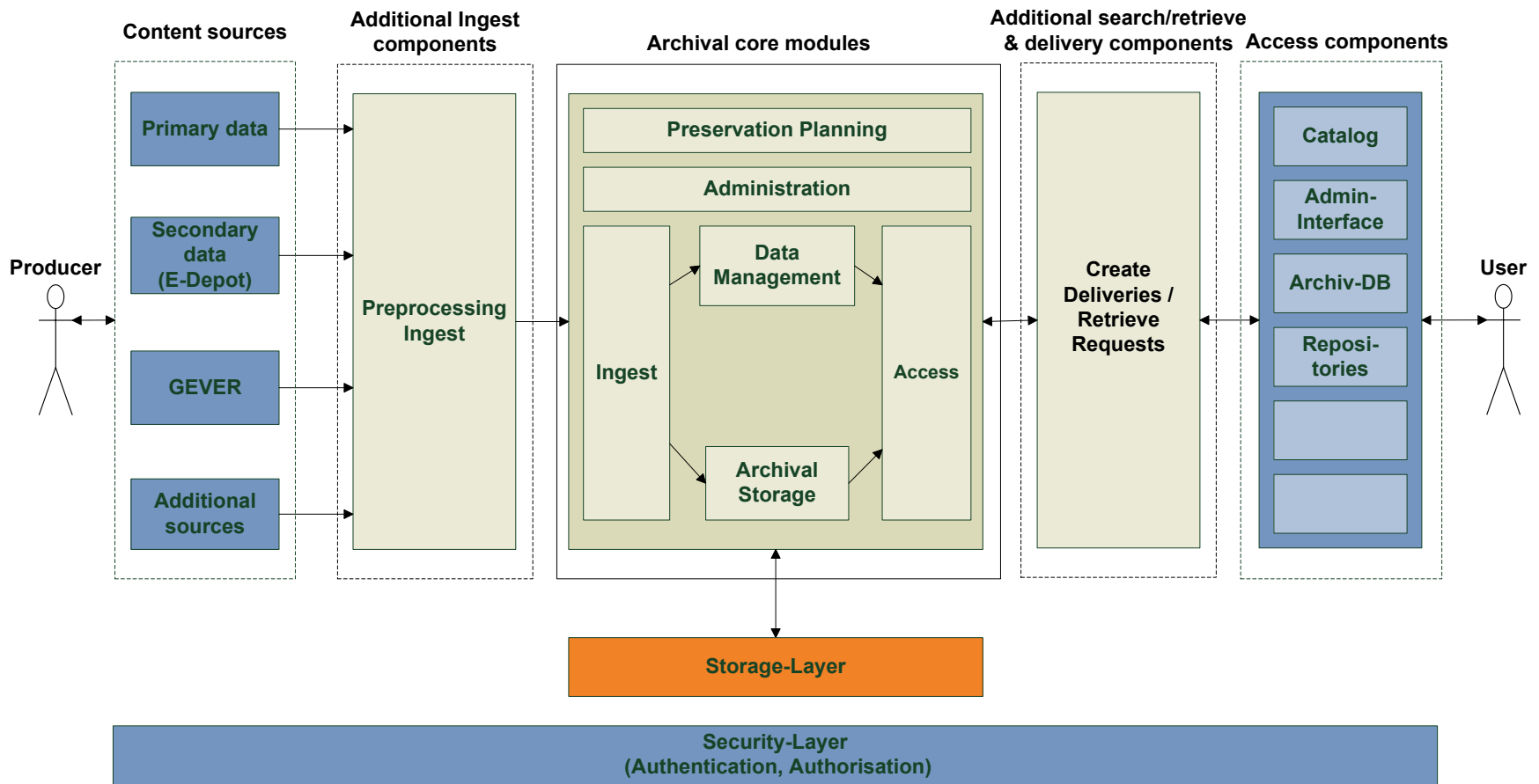
ETH-Bibliothek

- “**Digital Curation**”: Gestion de données et pérennisation
- Conservation de la **lisibilité technique** (Preservation) par monitoring et mesures actives (logiciel Rosetta)
- Si existant à la submission : Conservation de la **interprétabilité sémantique**, i.e. du context.

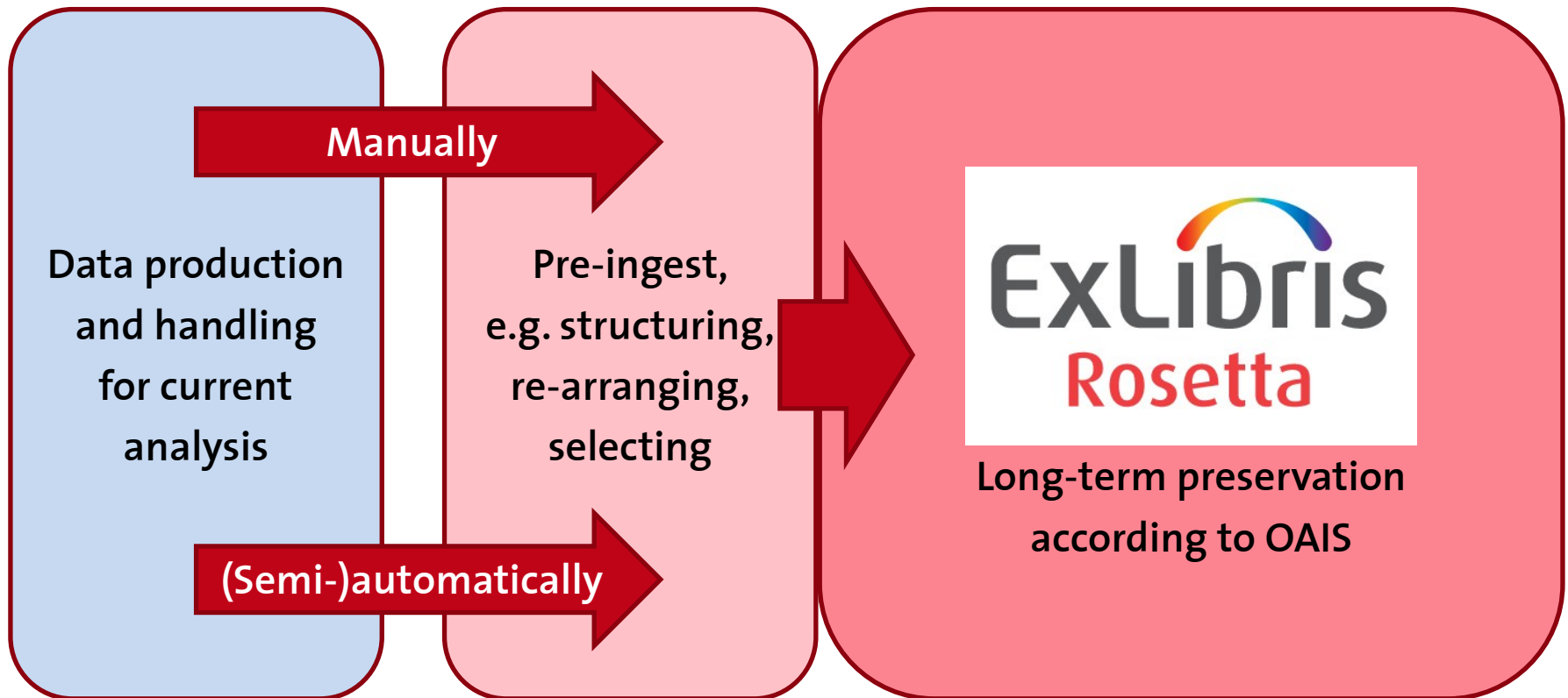
Services IT de l'ETH Zurich

- Opération et gestion de **l'infrastructure de la mémoire IT**.
- “**Bitstream preservation**”: stockage sûr, durée selon besoin
- Dépôt en accord avec exigences légales (**SAP-Module**)

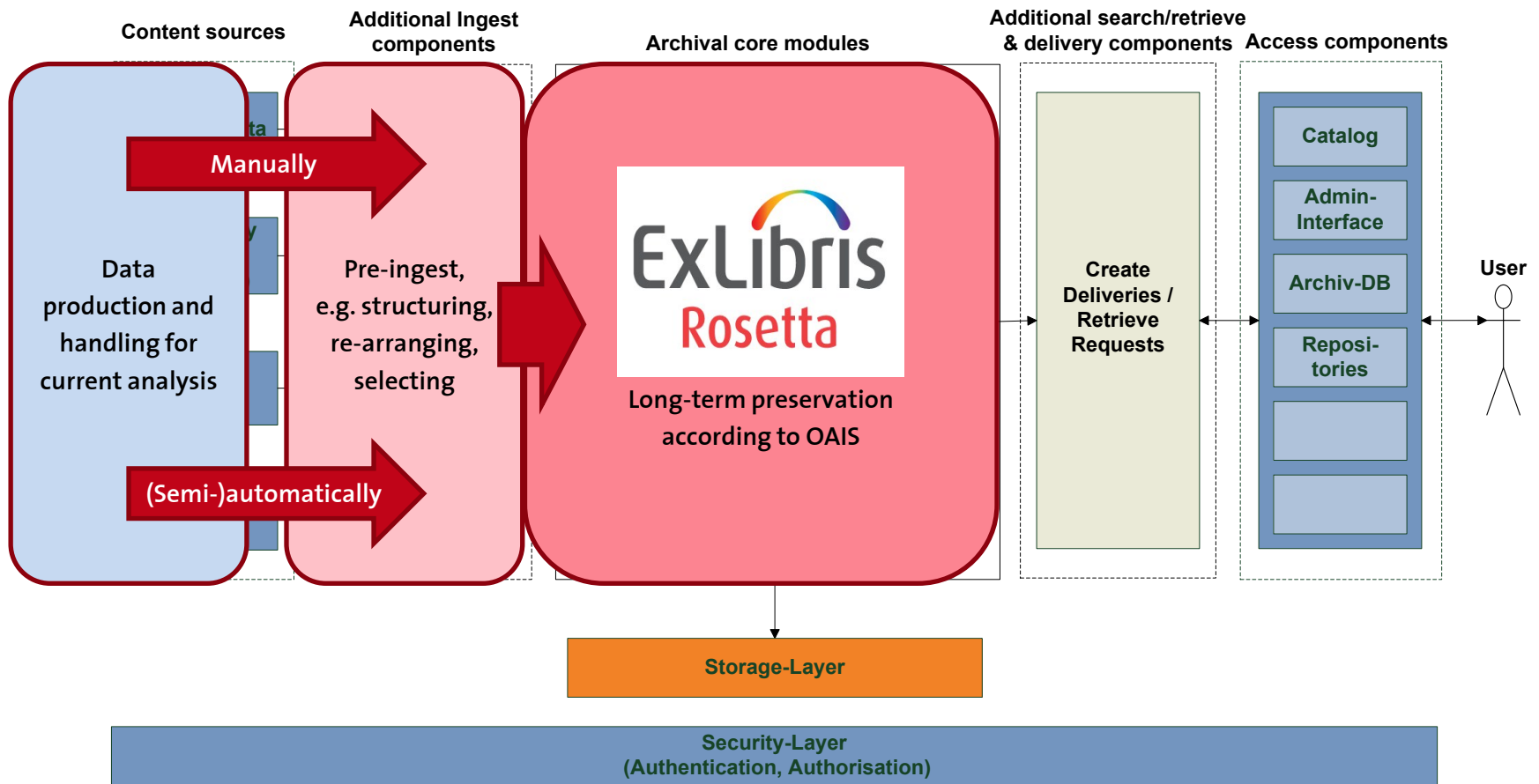
VISION – VUE TECHNIQUE



Abschlussbericht zur zweiten Phase „Pilot Langzeitarchivierung“, S. 23f; Aliesch, P. et al., 2007: Projekt „Pilot Langzeitarchivierung“. Intern.



ROSETTA DANS LA VISION



Abschlussbericht zur zweiten Phase „Pilot Langzeitarchivierung“, S. 23f; Aliesch, P. et al., 2007: Projekt „Pilot Langzeitarchivierung“. Intern.

PROCHAINES ÉTAPES

- Premiers «workflows» traitent **données de recherche** à être déposées en relation à une **publication** (2011/2012)
- Après, **données des dépôts structurés existants** doivent être importées (2012)
- **Partenariat du développement** en vue de la pérennisation des données de recherche convenu avec Ex Libris (2012)
- Entre autres il y a examiné d'installer de **fonctions de gestion de données** avant l'import dans Rosetta

BILAN INTERMÉDIAIRE

- Clarification importante avec tous les parties intéressées :
 - Ce qui couvre « l'archivage à long terme » ?
 - Quelles tâches sont implanté où ?
- S'applique surtout aux données de recherche :
«garbage in – garbage out»
- **Discussions intensives** avec chercheurs et autres parties intéressées sont nécessaires → effort
- Marche sur la corde raide:
Configuration adaptée à la demande et **individuelle vs. maniabilité et automatisation** de la solution globale

PERSPECTIVE DOMAINE DES EPF ?

- Intérêt pour l'échange d'information et **expériences**
- Actuellement **ni d'offre concrète, ni plans ou mandats**
- **Capacité limité**
- On essaye **d'éviter des activités qui compliqueraient** une coopération concrète ultérieure...
- ... p. ex. **Rosetta** a été choisi entre autres à cause de son qualité du logiciel **utilisable en consortium**

MERCI BEAUCOUP !

Questions ?

Dr. Matthias Töwe
ETH Zürich
ETH-Bibliothek
Leitung Digitaler Datenerhalt
Rämistrasse 101
8092 Zürich
Schweiz
matthias.toewe@library.ethz.ch
<http://www.library.ethz.ch>